

A LINEÁRIS DISZKRIMINANCIA-ANALÍZIS ÉS A LOGISZTIKUS REGRESSZIÓ MÓDSZERTANI KÉRDÉSEI

METHODOLOGICAL ISSUES OF THE LINEAR DISCRIMINANT ANALYSIS AND THE LOGISTIC REGRESSION

Sütő Dávid

Debreceni Egyetem, Gazdaságtudományi Kar
Vállalkozásfejlesztés Msc szak II. évfolyam

ÖSSZEFOGLALÁS

Ahhoz, hogy felmérhető legyen egy vállalkozás állapota, és az esetleges válsághelyzet, szükség van mérési, előrejelzési eszközökre. A belső működés folyamatos elemzése, a csődveszély felmérése, és csődelőrejelzési módszerek alkalmazása elengedhetetlen versenyképesség megőrzéséhez. Az elemzéseket kezdetben pénzügyi mutatószámokkal, később mutatószámrendszerek segítségével és statisztikai eszközökkel végezték az elemzők. Napjainkban a felgyorsult globalizálódott gazdaságban a módszerek ismerete és alkalmazása elengedhetetlen mind a vállalatok, mint vevők, szállítók, és egyéb érintettek értékelésekor.

Kulcsszavak: diszkriminancia-analízis, logisztikus regresszió, csődelőrejelzés

ABSTRACT

It, that can be measured let the undertaking's state be, and the potential crisis situation, need are onto measurement, forecast devices. The inner function the survey of his continuous analysis, the bankruptcy danger, and the application of bankruptcy forecast methods to the conservation of essential competitiveness. The analyses initially with financial indices, later mutatószámrendszerek the analysts made it with his help and statistical devices. Our Suns the accelerated one the knowledge of the methods and his application were becoming globalized in an economy essential all the companies, than their customers, their suppliers, and other affected his assessment.

BEVEZETÉS

A vállalatok tulajdonosai, vezetői (legyen az bármely vezetői szint) a vállalati életeciklusok egyes szakaszaiban különböző döntési helyzetekkel találják szemben magukat. Amennyiben a vezető nem foglalkozik kellő mértékben a számokkal, valamint a mögöttük levő összefüggésekkel és üzleti folyamatokkal, akkor az információhiányhoz, rossz döntések sorozatához és a vállalkozás instabil gazdasági helyzetbe kerüléséhez vezethet (BÓCSKEI et al, 2015). A gazdaságban kialakult és egyre inkább előtérbe kerülő bizonytalanságok, nehézségek következtében a vállalatirányítási munkában is egyre nagyobb szerepet kap a kockázatmérés és -menedzselés kérdésköre (ORBÁN és KISS, 2016). Erőteljesebb igény jelentkezik olyan megoldások iránt, amelyek segítségével a csődkockázat előre jelezhető. Egy jól irányított vállalkozás működésének hatékonyságát mérésrel is lehet/kell igazolni, és am mérési eredmények elemzésével lehet a további haladás irányát meghatározni. Nem (csak) attól hatékony egy vállalkozás, hogy azt marketing eszközökkel hirdeti, hanem attól is, hogy a szervezeten belül méri a hatékonyságot, és a hatékonyságnövekedést számokkal támasztják alá (FENYVES és DAJNOKI, 2015).

Gazdaság- és gazdálkodástudományi területen a csődelőrejelzés nem számít új területnek, hiszen már az 1900-as évek derekán is készültek elemzések és vizsgálatok a témában. Fejlődése viszont dinamikus, hiszen a 2000-es évek derekán járva elmondhatjuk, hogy széles tudományterületi és módszertani megközelítést, és töretlen módszertani fejlődést tudhat a háta mögött.

Az első tanulmányok mutatószám elemzést használtak a csődelőrejelzés módszerétül. Egyedi rátákat alkalmaztak, hogy összehasonlítsák a csődbement és sikeres vállalkozásokat. Ezek az

egyszerű módszerek fontos szerepet töltek be a későbbi csődmodellek kialakításánál, és adtak alapot a többváltozós csődmodellek kialakításához. (BELLOVARY et al., 2007). A mutatószámok használatakor előre definiált célt kell kitűzni. Ilyen cél a vállalatok csődelőrejelzése, amelyekhez fontos a megfelelő mutatószámok kiválasztása. A ráták önmagukban nem szolgáltatnak elegendő információt a következtetések levonásához, de kiinduló pontot jelentenek egy mutatószámrendszer összeállításához, az empirikus ellenőrzéshez és alkalmazáshoz (BEAVER, 1966).

EREDMÉNYEK

Fejlődéstörténeti alapok

Mutatószámrendszer, mint csődelőrejelzési eszköz

1930-tól 1965-ig a kutatók főként mutató számok segítségével próbálták elkülöníteni a csődbe ment és működő cégeket, és meghatározni a csődhelyzet okát. 1966-ban William H. Beaver fedtette le a csődmodellek alkotóelemeit képező mutatószámrendszer elmélet alapjait. 30 mutatószám segítségével hasonlított össze 79 csődbe ment és 79 működő vállalkozást, a vállalkozások adatai 38 különböző ágazatból kerültek kiválasztásra. Beaver empirikus kutatása során az adott mintára vonatkozóan következő eredményeket kapta:

- az Adózás utáni eredmény / Összes kötelezettség előjelzési képessége 92%,
- Adózás utáni eredmény / Árbevétel előjelzési képessége 91%,
- Adózás utáni eredmény / Nettó forgótőke előjelzési képessége 90%,
- Cash Flow / Összes kötelezettség előjelzési képessége 90%,
- Cash Flow / Összes eszköz előjelzési képessége 90%,
- Likviditási ráta előjelzési képessége 90% (BELLOVARY et al., 2007).

A kutatás célja az egyes mutatószámok egy évre történő előrejelző képességének vizsgálata volt. A pénzügyi mutatószámok indikátorként jeleznek előre, és egy-egy magyarázó változóként tekinthetők a csőd kialakulásakor. A kutatás szerint több mutatószám egyidejű figyelembe vételével szimultán döntéshozatal esetén magasabb előrejelzési képességet biztosít, mint az egyszerű mutatószámok külön-külön elemzése.

A mutatószámok – belső összefüggéseik révén – kölcsönösen hatnak egymásra, így sajátos komplex rendszert alkotnak. Az elemző felkészültsége határozza meg azt, hogy ebből az összefüggésrendszerből mit képes felszínre hozni.

A csődelőrejelzés módszertani alapjai, modellalkotás

A két leggyakrabban alkalmazott módszer csődelőrejelzési területen a diszkriminanciaanalízis, és logisztikus regresszió. A korábban kifejlesztett modellek, mint az Altman-, Springate-, Fulmer-modell alkalmazhatóak általánosan a magyar gazdasági viszonyok mellett, de egy ágazat vizsgálatához a magasabb pontosság érdekében indokolt specifikus modell kidolgozása.

Lineáris regresszió

A vizsgált jelenséget reprezentáló Y eredményváltozó, és az okok szerepét játszó X magyarázó változók kijelölése a modell specifikálásának első lépése. A modell struktúráját az $Y = X_1 + X_2 + \dots + X_p + \varepsilon$ formula írja le, ahol ε maradékváltozó, ami a modell által nem magyarázott, véletlen hatást fejezi ki. A véletlen változón keresztül ítélnél meg a modell és a valóság viszonya. A véletlen változó minden regressziós modell eleme, ezért a $(p + 1)$ változós modellre elegendő az $Y = X_1 + X_2 + \dots + X_p$ formában hivatkozni. A regressziós modell feltevése szerint az eredményváltozó feltételes várható értéke a magyarázó változó lineáris függvénye:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

ahol

$$\hat{Y} = \text{regresszió értéke}$$

$$\beta_0, \beta_1 \dots \beta_p = \text{regressziós paraméterek}$$

$$X_1, X_2 \dots X_p = \text{magyarázó változók.}$$

A regressziós függvény \hat{Y} értékét a későbbiekben tömören regresszióknak nevezzük. A regresszió értékének az ismeretében egy adott X feltétel mellett Y értéktől való ε eltérés = $Y - \hat{Y}$ (HAJDÚ, 2001).

A lineáris regresszió számításához használható az R Statistics 'lm' modulja, így a regressziós együtthatókkal együtt meghatározásra kerülnek az egyes magyarázó változók t-próba értékei (szignifikancia szintjei) is. Kiszámításra kerül továbbá a totális determinációs együttható (korrelációs együttható négyzete), illetve az F-próba értéke és szignifikancia szintje is. A próbák abban segítettek az elemzés során, hogy meg tudjuk állapítani, hogy egyes pénzügyi mutatók, mint magyarázó változók megbízhatóságát. A determinációs együttható megmutatja, milyen arányban tudjuk magyarázni a függő változó varianciáját a független változók varianciájával. Az F próba értéke pedig a regressziós függvény magyarázó hatását mutatja meg.

Diszkriminancia-analízis

A diszkriminancia-analízis egy statisztikai módszer, aminek a felhasználásával a pénzügyi mutatószámok – mint előredefiniált változók - segítségével különbséget tehetünk a csoporton belüli egyedek között. A diszkriminancia-függvény lehet lineáris és négyzetes, ennek megfelelően megkülönböztetünk lineáris és kvadrátikus diszkriminancia-analízist (BERG, 2005).

A többváltozós diszkriminancia-analízis egyidejűleg elemzi több mutatószám eloszlását, és olyan osztályozási szabályt állít fel, amely több súlyozott pénzügyi mutatószámot – ezek a modell független változói – tartalmaz, és ezeket egyetlen diszkriminancia-értékké fogja össze. Az alkalmazott mutatószámok kiválasztásának előfeltétele, hogy azok kevésbé korreláljanak egymással, különben a bevont újabb mutatószámok csak minimális mértékben járulnak hozzá a csoportképzés megbízhatóságának növekedéséhez. További feltétel, hogy a mutatószámok többdimenziós normális eloszlást mutassanak, valamint az, hogy az osztályok kovarianciamátrixai azonosak legyenek.

A vállalatok osztályozásához az egyes vállalatok éves beszámolóadataiból kiszámított pénzügyi mutatószám-értékeket kell behelyettesíteni a lineáris kombinációt képező diszkriminancia-függvénybe. A fizetőképés és a fizetésképtelen vállalatokat elválasztó diszkriminancia-értékkel összehasonlítva állapítható meg, hogy a cég a kettő közül melyik csoportba sorolható. A diszkriminancia-függvény általános alakja a következő:

$$Z = w_1 X_1 + w_2 X_2 + \dots + w_n X_n$$

ahol

Z : diszkriminancia-érték,

w_i : diszkriminancia-súlyok,

X_i : független változók (pénzügyi mutatók),

$i = 1, \dots, n$, ahol n a pénzügyi mutatók száma (VIRÁG et al., 2013).

A lineáris diszkriminancia-analízisre a lépésenkénti regresszió elvégzése után kerül sor. Az elemzés elvégzéséhez az R Statistics 'lda' modelljét kell alkalmazni. A diszkriminancia-függvény azokból a magyarázó X_i változókból tevődik össze, amelyek a lépésenkénti lineáris regresszió elvégzése után bent maradtak a modellben, és a legmagasabb volt a megbízhatóságuk. Másféleképpen fogalmazva, amelyekhez a legalacsonyabb AIC érték kapcsolódott.

A 'lda' modell által kapott Z értékeket felhasználva, és azokat sorba rendezve két vagy több csoportba sorolhatók a vállalkozások, a szükséges csoportosítási relációk meghatározása után a „cut off” értékek meghatározásával.

Logisztikus regresszió

A logisztikus regressziót, mint módszert nem elsősorban a regresszió miatt érdemes alkalmazni. A fő oka az, hogy a módszer kiválóan alkalmazható a pénzügyi mutatószámok, mint magyarázó változók és a bináris osztályok szerinti kategóriába sorolásra, és a besorolás valószínűségének becslése. Az eljárás a súlyozott független változókhoz a mintában szereplő vállalatok megfelelő elkülönítéséhez valószínűséggel kifejezett Z-értéket rendelnek hozzá. A logisztikus regresszió képlete az alábbi:

$$\begin{aligned} \Pr(\text{fizetőképes}) &= e^z / (1 + e^z) \\ \Pr(\text{fizetőképes}) &= e^{\beta_0 + \sum(\beta_j X_j)} / (1 + e^{\beta_0 + \sum(\beta_j X_j)}) \end{aligned}$$

ahol

β_j – regressziós paraméterek,
 X_j – független változók száma (pénzügyi mutatók)
 $j = 1, \dots, m$, ahol m a pénzügyi mutatók száma.

Az eljárás eredményeképpen kapott 0 és 1 közé eső szám valószínűségi értéként tekinthető (VIRÁG és KRISTÓF, 2005).

A modell alapja az 'odds', amely az előzőleg definiált két csoportba kerülés valószínűségeinek egymáshoz viszonyított arányát jelenti. Csődmmodellek esetében gyakorlatilag ez a túlélés és a csőd bekövetkezési esélyeinek hányadosát takarja. Képlettel kifejezve:

$$\text{odds}_x = P_{\text{túlélő}|X} / P_{\text{csőd}|X}$$

ahol x az adott megfigyelésre vonatkozó független változók ($x_1, x_2 \dots x_p$) vektora. Csődmmodellekre értelmezve pedig a vizsgált vállalatok mutatószámainak vektora (IMRE, 2008).

A logisztikus regresszió-modell felépítésének kulcskérdése a rendelkezésre álló változók számának megfelelő mértékű csökkentése. A logisztikus regresszióelemzés – szemben a diszkriminancia-analízissel – nem igényli a változók normális eloszlását, és az egyenlő variancia-kovariancia mátrixokat a két osztályban, problémát okozhat azonban a több változó együttes alkalmazásakor fennálló multikollinearitás, valamint a nem szignifikáns változók jelenléte. Továbbá fontos kiszűrni a kiugróan magas értékeket a modellezés megkezdése előtt (VIRÁG et al., 2013).

A kiugró értékek kiszűrése a minta előkészítése során kell megtörténnjen. A multikollinearitás problémájára, és a magyarázó változók számának csökkentésére a már fent ismertetett 'step' R Statistics modul nyújthat megoldást.

Lépésenkénti „stepwise” regresszió

Az eredményváltozóra legnagyobb hatással bíró magyarázó változók kiválasztásánál a többváltozós lépésenkénti lineáris és logisztikus regressziót kell alkalmazni. A lépésenkénti regresszióhoz mind a két esetben el kell végezni a normál logisztikus és lineáris regresszió–számítást, amelyhez az R Statistics 'lm' és 'glm' modulját kell használni. Ezt követően futtatni kell a 'step' modult, és a bemeneti adatokat az előzőleg futtatott 'lm' és 'glm' modul kimeneti adatai szolgáltatják. A 'step' modell lépésenként hagyja ki kellő magyarázó erővel nem rendelkező változókat, azzal a céllal, hogy leegyszerűsítse a modellt. A változó szelekció az AIC (Akaike Information Criterion) érték beállításával hajta végre. A módszer a következő összefüggés felhasználásával az AIC értéket minimalizálja:

$$AIC = 2 * [-\log(L(\beta_0, \beta_1, \dots, \beta_p, \sigma^2 | Y)) K]$$

ahol

p – a magyarázó változók száma

K – $(p + 2)$

β_i – az eredeti modell becslése során kapott regressziós együtthatók

σ^2 – az eredeti modell becslése során kapott becslült értékek varianciája

$\log(L(\dots))$ – log-likelihood függvény

A modell lényege, hogy az irreleváns magyarázó változók modellhez adása növeli a loglikelihood függvény értékét. Egyszerűen úgyis lehetne fogalmazni, hogy az AIC kisebb értéke jobb modellhez vezet (SHEATHER, 2009).

A 'step' modul futtatása után kapott változókkal le kell futtatni az 'lda' és 'glm' modult. Az így kapott eredmények alapján létrejön az adott mintára vonatkozó modell.

A modellek összehasonlítása

Előrejelzés-módszertani oldalról kimutatható, hogy a kevésbé szigorú alkalmazási feltételű módszerek sikeresebben beválnak a gyakorlatban, annak ellenére, hogy a művelőit gyakran pozitivisták kritikákkal sújtják. Az empirikus vizsgálatok során nem a győztes módszert kell keresni, hanem kombinálni kell őket, és szimultán módon alkalmazni mindegyiket a gyakorlati előrejelzések során (VIRÁG et al., 2013).

1. táblázat: Csődmodellek előnyeinek és hátrányainak összehasonlítása

Módszer	Előnyök	Hátrányok
Diszkriminancia analízis	<ul style="list-style-type: none"> - többváltozós modell - egyszerű felépítés 	<ul style="list-style-type: none"> - több korlátozó előfeltételezéssel és (linearitás, normalitás, előzetes csoportvalószínűségek) - csak sorrendi mérő - független változók súlya nem határozható meg
Regressziós modellek	<ul style="list-style-type: none"> - többváltozós modell - kvalitatív ismérveket is képes használni - független változók súlya kinyerhető - nem feltételez linearitást 	<ul style="list-style-type: none"> - érzékeny a független változók korrelációjára - érzékeny a kiugró vagy hiányzó értékekre - nagyobb mértékű nem-normális eloszlás torzíthatja - kis mintán nem hatékony

Forrás: IMRE, 2008

KÖVETKEZTETÉSEK, JAVASLATOK

Altman volt az első, aki többváltozós csődmodellt kifejlesztve függvényként tekintett a csődre, mint függő változóra, az egyes mutatószámokra pedig, mint magyarázó változókra. A cégek csoportosítását a függő változó előre definiált kategóriái alapján végezte el. Olyan magyarázó változókat alkalmazott, amelyek lineáris kombinációja a vállalatokat hatékonyan képes szétválasztani, és a függőváltozó kategóriákba besorolni. A fent említett modellhez gyártó-, termelő vállalkozások adatait használta. Tekintve, hogy a diszkriminancia-elemzés kis elemszámú mintákon is alkalmazható, és figyelembe véve a modell előrejelzési képességét, kijelenthető, hogy a megfelelő pénzügyi mutatószámokat kiválasztva a modell alkalmas volt szignifikánsan megkülönböztetni a csődbement és működő vállalkozásokat, és jó alapot teremtett a jövőbeni kutatásokhoz. A logit modell nem támaszt korlátozó feltételeket a változók normalitásával kapcsolat-

ban, mint a diszkriminancia-analízis, továbbá megengedi a diszkrét és folytonos változók használatát is. A modellben található β értékek jelzik az egyes mutatók fontosságát. Mivel a logit érték 0 és 1 közötti értéket vehet fel, azonnal valószínűségi változóvá alakítható. A két módszer közötti különbség továbbá, hogy a diszkriminancia-analízissel szemben a logit modell kis mintákon nem hatékony. A két modell tekintetében nem jelenthető ki egyértelműen, hogy melyik a hatékonyabb. A modellek egyidejű tesztelése során viszont kiválasztható a tevékenység és ágazat függvényében melyiket érdemes alkalmazni.

FELHASZNÁLT IRODALOM

- (1) Beaver, W.H. (1966): Financial Ratios as Predictors of Failure, *Empirical Research in Accounting: Selected Studies* 1966, p. 71 (2) Bellovary, J.M. – Giacomino, D.E. – Akers, M.D. (2007): A Review of Bankruptcy Prediction Studies: 1930 to Present, *Journal of Financial Education*, Vol. 33., pp. 2-4. (3) Berg, D. (2005): Bankruptcy Prediction by Generalized Additive Models, Department of Mathematics, University of Oslo, Norway, Statistical Research Report No. 1, p. 5. (4) Böcskei E. – Bács Z. – Fenyves V. – Tarnóczy T. (2015) Kockázati tényezők lehetséges előrejelzése, a gazdálkodás felelősségének kérdése a számviteli beszámolókból nyerhető adatok tükrében. *Controller Info* 2015/3: pp. 7-14. (5) Dékán T.né Orbán I. – Kiss Á. (2016): Az egyéb átfogó eredmény (OCI) kimutatásának értelmezése vezetői szemszögből. *Taylor Gazdálkodás- és szervezéstudományi folyóirat, A Virtuális Intézet Közép-Európa Kutatására Közleményei* 2016/1 pp. 46-50. (6) Fenyves V. – Dajnoki K. (2015): Controlling eszközök a humán erőforrás gazdálkodás területén. *Controller Info* 2015/3: pp. 68-73. (7) Imre B. (2008): Bázis II. definíciókon alapuló nemfizetés-előrejelzési modellek Magyarországi vállalati mintán 2002-2006). Ph. D. értekezés. Miskolci Egyetem. Vállalkozásmélet és gyakorlat Ph.D. program. pp. 47-56. (8) Sheather, J.S. (2009): *A modern approach to Regression with R*. Springer Science + Business Media, LLC 2009. p. 230. <http://link.springer.com/book/10.1007%2F978-0-387-09608-7> (letöltve: 2016. február 26.) (9) Virág M. – Kristóf T. – Fiáth A. – Varsányi J. (2013): Pénzügyi elemzés, csődelőrejelzés, válságkezelés. Bp. Kossuth Kiadó, pp. 55-230. (10) Virág M. – Kristóf T. (2005): Az első hazai csődmodell újraszámítása neurális háló segítségével. Bp. Közgazdasági Szemle. LII. évf. p. 150.